36-463/663: Multilevel & Hierarchical Models
Spring 2022
HW03 – Due Tue Feb 15, 11:59pm

- Please turn the homework in, as a single pdf, online in GradeScope using the link provided on the Assignment page on canvas.cmu.edu. Upload *one* file per person.
- Reading:
  - See the subfolder "reading" in the week05 folder on Canvas.
- There are two main exercises, each with "parts"...

Exercises

1. Return to the CD4 data from HW02, and consider the model

```
lmer.1 <- lmer(sqrt.CD4PCT ~ 1 + VISIT * treatmnt + (1 + VISIT | newpid),
               data=cd4)
```

Bring together all the residuals (except random effects residuals) that we have talked about, using library(HLMdiag):

```
> r.1 <- hlm_resid(lmer.1,level=1,include.ls=F)
> r.1s <- hlm_resid(lmer.1,level=1,include.ls=F,standardize=T)
> r.2 <- hlm_resid(lmer.1,level="newpid",include.ls=F)
> r.2s <- hlm_resid(lmer.1,level="newpid",include.ls=F,standardize=T)
> names(r.1); names(r.1s); names(r.2); names(r.2s)
```

Read the help file for `hlm_resid` if necessary to understand what all the compenents are. (You can create all these residuals "by hand" as I have shown in lecture slides and in the accompanying R files, but using HLMdiag is simpler and more reliable.)

(a) Make a facets plot of the marginal residuals, as a function of the marginal fitted values (use `scales="free_x"` if needed to make the plot legible). Explain in a sentence or two why a facets plot is not very useful for assessing model fit for this problem, whether we look at the first 12 children, or all 251 children).

(b) Make an ungrouped (that is, no facets) scatter plot of marginal residuals as a function of marginal fitted values, using the full data set (not just the first 12 children). Color the points for `treatmnt=1` kids and treatmnt=2 kids with different colors. Overlay a smooth fit (`geom_smooth` is the easist to use here). Explain, in a couple of sentences (optionally with some math):

- What is causing the dominant structure in this plot, and why that dominant structure is essentially irrelevant for checking the relationship between `sqrt.CD4PCT` and `VISIT`; and
- What in this plot makes you happy or unhappy about having a linear relationship between `sqrt.CD4PCT` and `VISIT` in the model.

(c) Make an ungrouped (that is, no facets) scatter plot of conditional residuals as a function of conditional fitted values, using the full data set (not just the first 12 children). Color the points for `treatmnt=1` kids and `treatmnt=2` kids with different colors. Overlay a smooth fit (`geom_smooth` is the easist to use here). Explain, in a couple of sentences (optionally with some math):

- Why the dominant structure in the marginal residuals is not also present in this plot of conditional residuals
- What might be causing the trend you see in this plot to be different from the trend in the plot of the marginal residuals.

(d) Use standardized residuals and standardized random effects estimates to assess the normality of ϵ_i, η_{0j} and η_{1j} in the fitted model, and to check for any outliers. Include qq plots for each, and accompany each plot with a sentence or two describing what is good or bad in that plot.

2. Continuing with the CD4 data...

(a) Make a table giving values of AIC, BIC, DIC, and cAIC (you compared two of these on the last assignment,using just AIC, BIC and DIC):

```
sqrt.CD4PCT ~ 1 + visage +            (1+visage|newpid)
sqrt.CD4PCT ~ 1 + visage + treatmnt + (1+visage|newpid)
sqrt.CD4PCT ~ 1 + visage * treatmnt + (1+visage|newpid)
sqrt.CD4PCT ~ 1 + VISIT  +            (1+VISIT|newpid)
sqrt.CD4PCT ~ 1 + VISIT  + treatmnt + (1+VISIT|newpid)
sqrt.CD4PCT ~ 1 + VISIT  * treatmnt + (1+VISIT|newpid)
```

Comment briefly on any similarities or differences in how the different criteria choose fixed effects.

(b) Now let's look at the random effects in the model

```
sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (1+VISIT|newpid)
```

Again, make a table giving values of AIC, BIC, DIC, and cAIC for the following models:

```
sqrt.CD4PCT ~ 1 + VISIT + treatmnt
sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (1|newpid)
sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (0+VISIT|newpid)
sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (1+VISIT|newpid)
```

You'll fit the first model with lm(), and the others with lmer(). Comment briefly on any similarities or differences in how the different criteria choose random effects. (Note that only BIC and cAIC have a strong theoretical justification here).

(c) Repeat part (b) but with the interaction `VISIT * treatmnt` in each model instead of just the main effects `VISIT + treatmnt`.